ACCELERATION OF LEARNING IN HYBRID NEURAL NETWORKS: A NOVEL APPROACH FOR THE DESIGN OF BRAIN CHAOSMAKERS

Berj L. Bardakjian and Alan Chiu

Institute of Biomaterials and Biomedical Engineering,
University of Toronto, Toronto, Ontario, Canada M5S 3G9
E-mail: berj@cbl.utoronto.ca

Abstract – Epileptic seizures correspond to episodes of increased rhythmicity of the normally chaotic activity in biological neural networks. We propose to use hybrid neural networks where artificial neural networks are used to control the biological neural networks by learning their different states. The learning is dramatically accelerated when using a conjugate gradient method in conjunction with the Fletcher-Reeves method of optimization.
Keywords – Neural networks, learning, chaosmakers

I. Introduction

In the healthy brain, the pattern of electrical activity is complex and chaotic. The onset of an epileptic seizure is characterized by rhythmic activity of lower complexity as shown in Figure 1. If neuronal dynamics are controlled to ensure that high complexity activity is maintained then we can achieve suppression of seizures. A brain *chaosmaker* would act to break the rhythmic electrical activity and thereby suppress epileptic seizures.

Considerable attention has been given by many groups [1], [2] to the possibility of exploiting the theory of nonlinear dynamical systems to control epilepsy . It is believed that breaking the rhythmic activity will provide a therapeutic intervention against epileptic seizures. The long-term objective of this research is to develop a device which is capable of learning the healthy chaotic dynamics of a small part of the brain and detect a change to a rhythmic pattern of activity. Once rhythmicity is detected the device would deliver an electrical stimulus which would restore chaotic activity. In order for such a chaosmaker to be relevant, its function must not rely on knowledge of the system equations, as these equations are not available for *biological neural networks (BNNs)*. The strategy must be based solely on information provided by a measured time series of the brain's electrical activity. Thus we begin our development of these strategies with the development of a time series model capable of learning chaotic dynamics using *artificial neural networks (ANNs)*. This would allow (a) detection of state transitions between chaotic and rhythmic states, and (b) delivering the appropriate stimuli to restore the chaotic state in such *hybrid neural networks (HNNs)* where BNNs are controllable by ANNs. Such a strategy preassumes that the *learning* of BNNs' states by the ANNs can be accomplished quickly to allow for the required *adaptivity.* Traditional learning paradigms were found to be too slow to be feasible. Hence, we have identified the need for *acceleration* of the learning process.

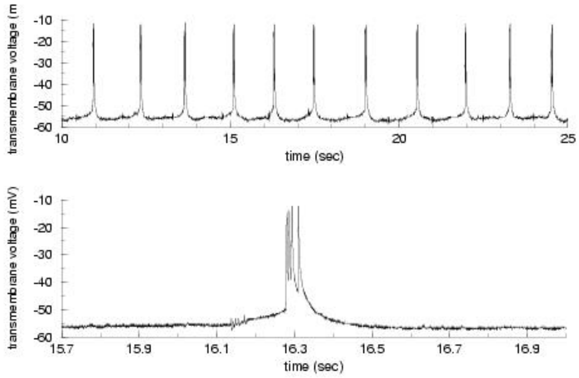

Fig. 1. The transmembrane voltage of a CA3 neuron in the rat hippocampal slice under conditions of zero calcium in the bathing solution (top trace), with a typical burst from the transmembrane voltage recording (bottom trace).

In this paper, we compare the efficiency of learning in feedforward artificial neural networks using three different optimization algorithms: *(i) Gradient Descent with Momentum (GDM), (ii) Conjugate Gradient Fletcher-Powell (CG/FP) and (iii) Modified Conjugate Gradient Fletcher-Reeves (CG/FR)* [3]. Most nonlinear optimization strategies utilize nonlinear approaches for finding search directions (such as GDM, CG/FP and CG/FR) and linear approaches for finding step sizes (*a.k.a. learning rates*) along the search directions (such as GDM and CG/FP). As an alternative technique to linear line search methods, the step size of the *Modified* (CG/FR) method is not selected based on criteria to reach a minima along each search direction. Rather, it is chosen to satisfy an inequality constraint based on the criteria that the eigenvalues of the estimated inverse Hessian matrix must tend monotonically to those of the actual inverse Hessian matrix. Simulations are performed on two classification problems, digit recognition and signal classification network. With exit condition fixed for training, the simulations show that drastic acceleration of learning can be obtained while maintaining good network generalization ability when the CG/FR optimization tech-

Report Documentation Page

Report Date	Report Type	Dates Covered (from... to)
25OCT2001	N/A	-

Title and Subtitle Acceleration of Learning in Hybrid Neural Networks: A Novel Approach for the Design of Brain Chaosmakers	**Contract Number**
	Grant Number
	Program Element Number
Author(s)	**Project Number**
	Task Number
	Work Unit Number
Performing Organization Name(s) and Address(es) Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, Ontario, Canada M5S 3G9	**Performing Organization Report Number**
Sponsoring/Monitoring Agency Name(s) and Address(es) US Army Research, Development & Standardization Group (UK) PSC 802 Box 15 FPO AE 09499-1500	**Sponsor/Monitor's Acronym(s)**
	Sponsor/Monitor's Report Number(s)

Distribution/Availability Statement
Approved for public release, distribution unlimited

Supplementary Notes
Papers from the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, October 25-28, 2001, held in Istanbul, Turkey. See also ADM001351 for entire conference on cd-rom., The original document contains color images.

Abstract

Subject Terms

Report Classification unclassified	**Classification of this page** unclassified
Classification of Abstract unclassified	**Limitation of Abstract** UU

Number of Pages
4

nique is used.

II. Methods

Learning in artificial neural networks can be formulated as an optimization problem where the objective is to minimize a mean-squared error (MSE) function of the form

$$E(\mathbf{w}) = \frac{1}{N}\sum_{i=1}^{N}(y(\mathbf{w}, x_i) - d_i)^2 \tag{1}$$

defined on a training set

$$\tau = \{x_i, d_i\},\ for\ i = 1, 2, ..., N \tag{2}$$

where $\mathbf{w}$ is the vector of synaptic weights, N is the size of the training set, x_i is the input, d_i is the desired output, and $y(\mathbf{w}, x_i)$ is the actual output.

The objective of the learning is to find a set of weights $\mathbf{w}^*$ which minimizes the objective function $E(\mathbf{w})$ or reduces it to a satisfactory level. The standard approach for solving such problems is to regard the objective function as an *n-dimensional* surface in the weights space and to employ algorithms which repeatedly take steps along this surface which are in the downhill direction until a minimum is reached. In gradient-based optimization methods, such search directions δ_k are functions of the gradient of the objective function $E(\mathbf{w})$ with respect to the synaptic weights. Such methods have a learning rule for iterative updating of the synaptic weights vector of the form

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \theta_k \delta_k \tag{3}$$

where θ_k is the learning rate (or step size) and δ_k is the direction of search at the kth iteration.

The GDM and CG/FP methods utilize a full linear search for finding the step size θ_k which minimizes $E(\mathbf{w})$ along the search direction δ_k. On the other hand, the *Modified* Conjugate Gradient Fletcher-Reeves (CG/FR) does not utilize a full linear search to find θ_k, and instead it is chosen to satisfy an inequality constraint based on the criteria that the eigenvalues of the estimated inverse Hessian matrix must tend monotonically to those of the actual inverse Hessian matrix.

III. Results

The three learning paradigms were tested using a feedforward neural network (with 10690 synaptic weights, 256 input neural units and one layer of 40 hidden neural units) to perform a standard digit recognition task where the desired outputs d_i are as shown in Figure 2. The three learning paradigms were started from the same initial starting point with the $MSE = 1$ at the zeroth iteration. The learning was specified to be completed when a target level of $MSE = 0.01$ was reached. The target value of the mean squared error for the three learning paradigms was reached after {7121, 355 and 66} iterations for the {GDM, CG/FP and CG/FR}, respectively. This suggests a dramatic acceleration of learning when using the CG/FR paradigm as shown in Figure 3. The synaptic weights are depicted in Figure 4.

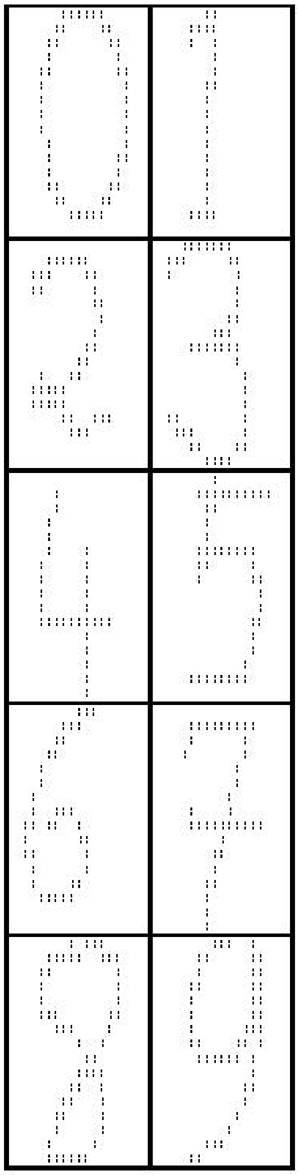

Fig. 2. A digit recognition task used as an example to compare the speed of learning using the three learning paradigms.

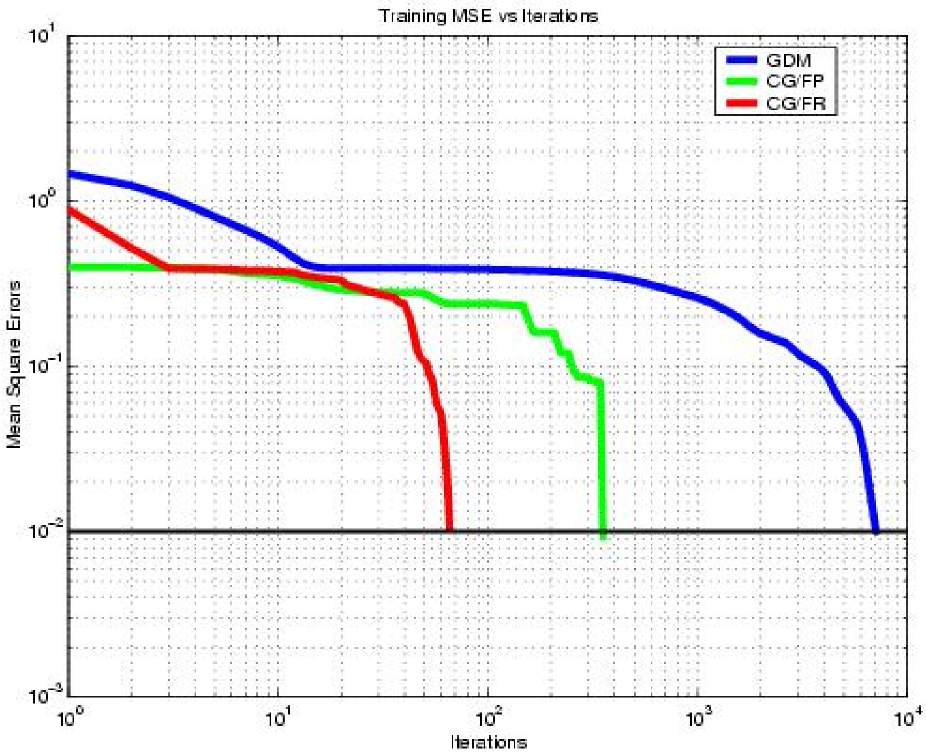

Fig. 3. The comparison of the three methods of optimization used in the learning phase. The specified mean squared error (MSE = 0.01) for the three learning paradigms is reached after {7121, 355 and 66} iterations for the {GDM, CG/FP and CG/FR}, respectively. Note that at the zeroth iteration, the three methods were started at an initial MSE=1, but the three paradigms had different MSEs after the first iteration.

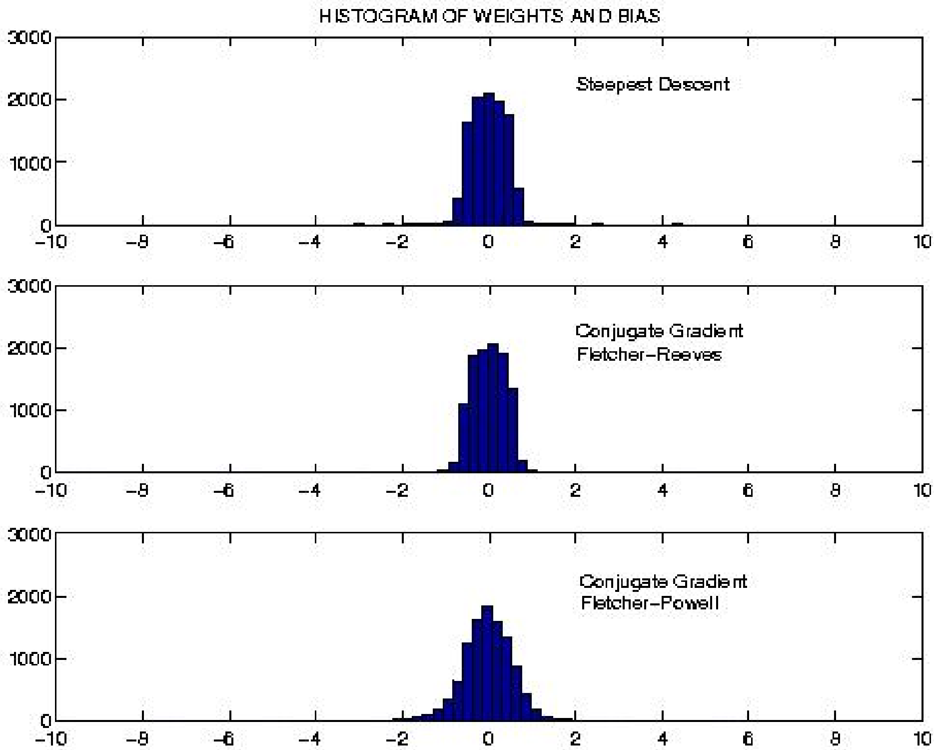

Fig. 4. The histograms of the synaptic weights using the three learning paradigms.

IV. Discussion

One perspective on the problem of maintaining chaos in biological systems is that of *Schiff et al.* [2], who attempted to *anticontrol* chaos in the rat hippocampal slice. They used linear regression to fit eigenvectors and find an unstable saddle point in the first return map of the interspike interval. Their efforts to maintain chaos in the neuronal tissue was limited to eliciting action potentials, through stimulation, such that the interspike interval landed off the eigenvectors. Their perspective was one of increasing the variability in the neuronal activity over that observed by their model.

In our approach, the concept of targeting introduced by Shinbrot *et al.* [4] is used. Once stabilization is detected, the ANN is used to iterate a population of trajectories originating within an ϵ neighbourhood of the current position of the BNN's state in the state space. These iterations form an approximation to the chaotic BNN dynamics.

The control algorithm waits until a placement of the output causes the *observed System* dynamics to fall directly onto the ANNs approximation of the BNN's unstable manifold. This action causes a temporary return to chaotic activity by perturbing the BNN's dynamics out of the stable region of state space and into the chaotic region.

Comparing our approach to that of *Schiff et al.*, we see that both strategies employ model estimation from time series and make perturbations directly to the system variable. However, unlike our strategy, Schiff et al. do not consider the system as having gone through a transition. They focus their efforts on learning the low dimensional epileptic dynamics, whereas we focus on learning the higher complexity dynamics of the healthy activity. Our advantage is that once a transition to rhythmicity is detected, the initiation of a control action may begin immediately without a lengthy learning stage during the seizure activity. The approach of Schiff et al. has the advantage of learning the less complex dynamics which could be a considerably easier task than learning the high complexity dynamics. Nevertheless, the use of the accelerated learning paradigm described in this paper, may render the learning of the higher complexity dynamics, feasible.

The choice of time series model is a second point of comparison between our approach and that of Schiff et al.. The algorithm of Schiff et al. learn the local dynamics of a unstable periodic orbit (UPO) by fitting linear eigenvectors in the two dimensional first return map. The goal of our ANN model is to learn the global dynamics of the chaotic system representable in any dimension. The literature regarding the detection of nonlinearities in neuronal activity largely suggests that embedding the interspike interval in a two dimensional state space will result in a significant number of false nearest neighbours, complicating the learning of deterministic dynamics. In the applications of chaos control algorithms, the local dynamics may be well described by the two dimensional linear model; however in instances of *anticontrol*, the dynamics of interest are those of the greater chaotic attractor and not the local dynamics.

V. Conclusion

In conclusion, we have developed a novel approach to a potential therapy for epileptic disorders. The strategy is to learn the global dynamics of the healthy chaotic systems and their transition to rhythmicity in BNNs using ANNs. Such learning is accelerated significantly using a conjugate gradient method without performing a full linear search for the minimum, along the direction of search, in each iteration. Whenever rhythmicity in the BNN is detected, the control strategy employs the ANN to estimate the unstable manifold of the rhythmic orbit upon which it will place the state vector. This has the effect of restoring the BNN to its chaotic state.

VI. Acknowledgments

This work has been supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

VII. References

[1] Freeman, Walter J., "Tutorial on Neurology: From Single Neurons to Brain Chaos," *Int. J. of Bifurcation and Chaos.*, **2**(3), 451-482, 1992.

[2] Schiff, S. J., Jerger, K., Duong, D. H., Chang, T., Spano, M. L. and Ditto, W. L. "Controlling Chaos in the Brain," *Nature*, **370**, 615-620, 1994.

[3] Fletcher, R., *Practical Methods of Optimization.* Chapters 3 & 4, John Wiley, 1990.

[4] Shinbrot, T., Grebogi, C., Ott, E. and Yorke, J. A., "Using Small Perturbations To Control Chaos," *Nature*, **363**, 411-417, 1993.